Bolt Beranek and Newman Inc.

Report No. 3520

Command and Control Related Computer Technology
Part I. Packet Radio
Part II. Speech Compression and Evaluation

March 1977

Submitted to:
Defense Advanced Research Project Agency

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: BBN REPORT NO. 3520	2. GOVT ACCESSION NO.: AD-A108 325
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): COMMAND AND CONTROL RELATED COMPUTER TECHNOLOGY. Part 1. Part 2.	5. TYPE OF REPORT & PERIOD COVERED: Quarterly Progress Report 1 Dec. 76 - 28 Feb. 77
	6. PERFORMING ORG. REPORT NUMBER: BBN Report No. 3520
7. AUTHOR(s): M. Beeler, J. Burchfiel, R. S. Nickerson, R. Viswanathan, J. Makhoul, A. W. F. Huggins	8. CONTRACT OR GRANT NUMBER(s): MDA903-75-C-0180 ARPA Order-2935
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Bolt Beranek and Newman Inc. 50 Moulton Street, Cambridge, Massachusetts 02138	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency Information Processing Technology Office 1400 Wilson Boulevard, ARlington, VA 22209	12. REPORT DATE
	13. NUMBER OF PAGES: 64
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited. It may be released to the Clearinghouse; Department of Commerce for sale to the general public.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was supported by the Defense Advanced Research Projects Agency under ARPA Order No. 2935.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

packet radio, computer communications, PDP-11, TCP, station gateway, ELF, BCPL, cross-radio debugging, speech compression, vocoder, linear prediction, variable data rate compression, speech-quality evaluation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This document describes progress on (1) the development of a packet radio network, (2) speech compression and evaluation. Activities reported under (1) includes work on Station Software and Internetworking Research and Development; under (2) comparative informal evaluation of our perceptual-model-based variable frame rate scheme for transmitting LPC data with two other schemes under a number of vocoder conditions; development and testing of a new scheme for variable rate transmission of LPC gain parameter;

COMPLETION OF THE INITIAL ANALYSIS OF OUR RECENT FACTORIAL SUBJECTIVE SPEECH QUALITY STUDY.

BBN Report No. 3520 March 1977

COMMAND AND CONTROL RELATED COMPUTER TECHNOLOGY

Part I. Packet Radio
Part II. Speech Compression and Evaluation

Quarterly Progress Report No. 9

1 December 1976 to 28 February 1977

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of the Defense Advanced Research Projects Agency or the United States Government.

This research was supported by the Defense Advanced Research Projects Agency under ARPA Order No. 2935 Contract No. MDA903-75-C-0180.

Distribution of this document is unlimited. It may be released to the Clearinghouse Department of Commerce for sale to the general public.

I. INTRODUCTION

The packet radio project relies heavily on station software for a variety of control, coordination and monitoring functions. The role of BBN in developing this software is to specify, design, implement and deliver programs which implement these functions.

At the close of the previous quarter we were about to deliver a gateway software package, and were awaiting arrival of CAP3 packet radio unit software from Collins Radio. The gateway was successfully demonstrated and delivered to University College of London; and CAP3 arrival permitted significant progress in station software development this quarter, including delivery of a new version to SRI.

UCL gateway doesn't depend on PR CAP! Non sequitur.

Besides completion of these two tasks, whose conclusion was anticipated but not yet reached at the close of the last quarter, continuing effort on the measurement process has brought it well along the road of implementation. Negotiations with other contractors, UCLA in particular, have been especially fruitful in solidifying the specification of file entries to be made by the measurement process. This and related measurement file issues have been documented and released to the working group.

its 28 April now and still no meas. process.

Further work this quarter has been expended to improve the ELF system and the cross-network debugger. Some bugs in ELF have been fixed, but the major ELF effort has been the design,

documentation and publication of a plan for enhancements to be installed in the coming quarter. These modifications will permit either disk or cross-network loading, thus reuniting the temporarily divergent paths BBN and SRI ELF systems had taken.

Additionally, this quarter we found a new employee to join the BBN packet radio group. Her extensive PDP-11 experience will be of great benefit to our efforts in this area, and we expect her to take on implementation of the station's information/directory service process in the coming quarter.

The specific accomplishments of the quarter are discussed in detail in the sections below.

II. MEETINGS, TRIPS AND PUBLICATIONS

One meeting was held this quarter, and two major publications were released by BBN. In addition, delivery of the gateway software at UCL was the occasion for a trip by a BBN representative. While in London, she also attended a meeting on internetworking issues, such as the design and function of gateways. Since the packet radio station also serves a gateway role, this exchange of information through BBN is important in maintaining compatibility, efficiency and maximum capability among gateways.

The February 23-25 meeting at SRI centered on packet radio implementation issues. BBN personnel exchanged both data on present performance, and development plans for the coming months, with other implementors. Of particular interest was resolution on the use of Distress ROPs, or DROPs, by mobile terminals. The response of present station and network algorithms to a mobile terminal's loss of RF connectivity along its assigned route is too slow. SRI expressed concern that a better method be working in time for demonstrations scheduled for this spring. Fortunately, BBN had considered this problem and had already presented its solution, DROPs, in informal messages during this quarter. Thus, although an open airing was given to various alternative schemes proposed by other contractors, the DROP plan provided a well designed standard for comparison and was ultimately selected by the group. Decisive action on this issue,

together with prior BBN consideration of DROP implementation issues, gives strong assurance of successful operation in the spring demonstrations.

ATR should address mobile handoff tech issues not just chronology.

Other meeting issues relevant to the station development effort at BBN are:

overstatement.

* A BBN representative will be a member of an advisory committee to help Collins design and implement the next generation of packet radio digital units.

* BBN is to participate with other Boston area groups in use of a Boston area packet radio network, as hardware becomes available.

* A conceptual framework for internet measurement control and data collection was presented by BBN and generally accepted. The operation of the near term packet radio net measurements will differ from this schema mostly in use of raw ARPA net packets for data delivery to UCLA.

* Station operator terminal control facilities are needed as soon as possible. BBN will design and implement a module to do this in the coming quarter.

* BBN is to further investigate the station's bandwidth in processing normal ROP packets.

* BBN is to study uses of a programmable set of display lights to be implemented on the next generation of packet radio digital units.

* BBN is to deliver the first version of the measurement process during the next quarter.

The documents published by BBN this quarter are:

* PRTN 174-revision 2, "Packet Radio Network Station Labeling Process." This revision describes recent improvements to the label process. In particular, PRs are now relabeled at a lower hierarchy level when the labeler finds this feasible. The labeler documented is also updated for compatibility with CAP3 PR software.

is this always a good strategy when PTP is available?

* PRTN 212, "Specification of Measurement File Entries." This note defines and describes the entries made in the measurement file, by the measurement process, as a result of conditions

arising in, decisions made by, and network status observed by, station software. It provides a structured framework within which the various entries are assigned clean, logical formats. It will be expanded in the future to include cumulative statistics packets and pickup packets as they are defined by UCLA and Collins.

* PRTN 215, "Measurement File Delivery Specification." Questions of transfer protocol and intermediate storage on the station disk had clouded the issue of measurement delivery from the station to UCLA. The viable alternatives are presented and weighed. A flexible means is chosen and its operation defined. This discussion is preceded by presentation of a model for network and internetwork experiments, which sets the context for the remainder of the paper.

* Informal consultation was provided in significant quantity on three issues. First, we presented the BBN plan for DROPs emitted by mobile terminals, from which the resolution at the implementors' meeting grew. Second, preliminary specifications of the measurement file were furnished as part of our negotiations with UCLA, which resulted in PRTN 212, and which are responsible for the acceptance of PRTN 212. Third, we provided initial investigation and discussion of the bandwidth of the station forwarder. Measurements suggested by us and carried out by SRI largely exonerated our forwarder design and helped Collins in their decision to modify PR software. By reducing blockage of the station-to-PR interface, packet loss will be brought under control.

III. STATION SOFTWARE

A. Enhancements

In mid-December Collins released a new version, CAP3, of PR software. This required a new version of station software to run with it. We delivered the new software, containing CAP3 changes as well as other enhancements, at the end of the quarter. The delivery was somewhat delayed because SRI's PRs required hardware modifications in order to run CAP3. Testing of the station at SRI and transfer of files were done over the ARPANET from BBN. We also carried out a remote demonstration (performed by SRI personnel as instructed by us) to familiarize SRI personnel with changes in the station. In addition to software, the delivery included updates to the station operator's manual.

All four station modules -- the gateway, debug process, connection process, and control process -- were modified to accommodate the CAP3 packet header, which is one word longer than the CAP2 header. Additional changes were as follows.

Initialization was added to the gateway to permit in-core restarts, as requested by SRI. station?

The "LO" (load overlay) command was removed from the debug process, rather than incorporate the totally different PR software overlays of CAP3. The LO facility had been found by SRI to be of litt'e use.

The connection process was modified to reject any packet not containing a full-sized header. Originally it was possible for headers to vary in length; up to three of the four route words could be omitted if the route was short. However, all contractors agreed that variable-sized headers were of dubious utility and complicated packet processing, so they have been eliminated.

As a result of forwarding tests run by SRI and Collins, the number of buffers in the connection process available for forwarding was increased from two to eight. Under heavy traffic loads, the station PR was blocking the station, sending it packets but not always willing to receive them back, thus causing packets to be dropped in the station. Increasing the number of buffers helped reduce packet loss while the station waited for the station PR to accept packets.

Most of the station changes were in the control process. First, the content and format of ROP and Label packets were changed in CAP3, so the control process was changed to conform to the new packet definitions. Second, several changes were made at SRI's request to improve the operator interface and aid in network diagnosis. The connectivity report was revised to show bidirectional links more clearly and to show each labeled PR's level; printout of the header and some text in the error messages for bad ROPs/TOPs was added; and the default was changed to not force operator definition of route formats during initialization.

Third, the control process was enhanced to take advantage of two new capabilities provided by CAP3 -- unlabeling and incremental routing. In CAP3, a label packet with certain text reinitializes the PR. The station unlabels a PR which reports incorrect labeling (labeling the station didn't give it) if it cannot relabel the PR. This may aid the station in selecting labeling for the PR, since the PR's connectivity will be learned faster from its unlabeled ROPs. In any case, the unlabeling prevents the PR from operating with a label which may have the wrong format or duplicate the labeling of another PR.

The incremental routing capability of CAP3 allows a labeled PR to have an incomplete route to the station -- i.e. some or all route labels after the first may be unspecified. Any PR transmitting an inbound packet whose route is exhausted fills in the next hop of the route from its own route. The station can be set (using new commands) to assign either full routes or just the first hop, and indicates in its display of routing which parts the PRs were told. If a PR's route is changed while its level remains the same (as for example might happen in rerouting around a failed repeater at the next level), then the station propagates the change in its tables to the routes of outer PRs which route through the changed PR but were not told that part of their route; it does not have to relabel those PRs. Thus the use of incremental routing can reduce relabeling and allow smoother adjustment of routing.

Finally, two labeling improvements were made that do not relate directly to CAP3. Previously, the station only relabeled a PR if it thought something was wrong with its route. Now it also relabels a PR if it can place it at a lower level. It notices this possibility when it receives a ROP from the PR forwarded through a lower-level PR than the one the PR currently routes through.

On receiving a ROP, the station processes the PR that originated the ROP -- checking its labeling, entering it in its tables if it's a new PR, etc. In addition, it now does some processing of the PR that forwarded the ROP. The PR may be entered in the station tables, and may be labeled or unlabeled. This greatly improves initialization of an already-labeled net, which now takes only seconds; previously it could take minutes.

The behavior of the current control process is described in PRTN 174, revision 2, which was issued along with the software delivery.

B. Maintenance

A bug which caused a spurious error message to be typed when an acknowledgement for a TOP (Terminal-On-Packet) was sent by the control process was found and fixed. This feature had not been thoroughly tested prior to delivery last quarter because no terminals had been upgraded to emit TOPs.

Two problems in the ELF operating system were remedied this quarter. One involved a kernel stack overflow bug in the SRI I/O speedup changes. The other was a reappearance of a bug in the kernel code to maintain the CPU time used by each process. This bug had been found by BBN some months before, but reappeared ~~was not fixed~~ in the latest ELF release from SRI.

C. Measurement Development

Last quarter we implemented the collection of label and connectivity data from the control process by the measurement process. The label data was changed this quarter to accommodate the assignment of partial routes and the automatic propagation of route changes in incremental routing. Measurement development continued with the specification of cumulative statistics to be reported by the control and connection processes, and the implementation of the cumstats and their collection in the control and measurement processes. Cumstats in the connection process will be implemented next quarter. A detailed specification of all measurement entries defined so far -- namely those reporting station cumstats, labeling, and connectivity -- was published in PRTN 212, "Specification of Measurement File Entries."

Cumstats from PRs will be controlled and collected by the station over SPP connections. The station opens a connection and sets the PR's cumstat parameters, then the PR automatically sends the cumstats over the connection at the specified interval. We

negotiated with Collins a change to the handling of SPP connections in the PR, so that the PR would abort a connection if it failed to receive an acknowledgement for a packet it originated after some number of retransmissions. Without this change, the PR would simply go on to the next packet, and would never terminate a connection on its own initiative. This could have caused problems if the station had to be reinitialized while a measurement run was in progress, since the PRs would continue sending cumstats over non-existent connections, unbeknownst to the station.

IV. INTERNETWORKING RESEARCH AND DEVELOPMENT

A. Timestamping in Gateways

This quarter, we attended the Gateway and Satellite Network meetings held at University College London (UCL) in early December. Immediately prior to that meeting, we delivered the gateway software to UCL and demonstrated use of the gateway and the message generator and statistics gathering programs. We spent several days helping the people at UCL to interface their software to the gateway program and to familiarize themselves with the use of the cross-net debugger.

At the time of the gateway delivery to UCL, a general timestamping mechanism had been designed and implemented in the gateway. We had intended to debug this code and deliver the new version of the gateway to UCL shortly after the initial delivery. However, at the Satellite Network meeting at UCL, it was noted that severe buffering restrictions in the SIMPs (Satellite IMPs) would prevent use of a general timestamping scheme. In light of this, a special timestamping scheme was designed for use in gateways on the Satellite Network. This new timestamping scheme was implemented in the gateway and debugged by sending packets between the UCL and BBN gateways. The new version of the gateway containing the timestamping code was delivered to UCL in February.

The principal differences between the timestamping scheme implemented for the Satellite Network and the more general scheme proposed earlier are that the new timestamps are shorter, thus allowing a smaller range of values, and that an identifier of the SIMP or gateway which inserted the timestamp is no longer inserted in the packet. This latter change restricts use of timestamping to packets whose path through a set of SIMPs and gateways is known and unchangeable. This is not a severe restriction for the initial set of measurement tests as the routing through gateways and SIMPs is fixed.

B. XNET Improvements

A number of improvements were made this quarter to XNET, the cross-net debugger for PDP-11s. These improvements overcame problems which had been encountered in operation with unreliable networks or network interfaces and adapted XNET for better operation under the DEC TOPS20 operating system.

In order to improve reliability, checksums were added to messages which permit the discovery of message corruption in passing through the networks or network interfaces. The checksum was added in a way which makes its use optional. This was necessary since the internet bootstrap program could not accommodate the additional code required to supply checksums. The checksum was added as an additional word of text in the message so that the header format would not have to be changed. The presence of the checksum is detected by the data length field

(in the internet header) being larger than necessary to hold the actual data of the message. The checksum used is a simple 1's complement sum of the 16-bit words in the message.

Adding checksums to XNET required changes to the PDP-11 debugger module to check and generate them and to the TENEX/TOPS20 XNET program to do likewise. It also became necessary to modify the internet bootstrap program due to the presence of a bug which caused the data length field to be incorrectly set. This was of no consequence in previous operation because it was not examined by anything. But, of course, it caused incorrect operation with checksums since the detection of the presence of the checksum depends on the data length field. It would also have caused problems if such messages passed through gateways which used that field to supply the message length.

The addition of retransmission to XNET was required to permit operation with lossy networks. This need was amplified by the addition of checksums since data errors would otherwise cause hangups rather than errors. Previous modification to the PDP-11 debugger made the addition of retransmission relatively easy. These modifications were designed to gracefully handle duplicate messages. The XNET protocol does not directly provide for duplication detection; there are no sequence numbers per se. However, most commands cause no problems if repeated. Originally, certain repeated messages would cause a negative

response, e.g., setting a breakpoint which was already set would reply "can't". The modifications made earlier altered the response to indicate success although no action was actually taken.

The only program modifications required to implement retransmission were in the XNET program itself. The principal change was to replace the infinite wait for an acknowledgement with a finite wait after which the program loops back and transmits the message again.

Selecting the algorithm for determining the retransmission interval provided an opportunity for a certain amount of experimentation. The final algorithm is to maintain a short term average round-trip delay and a short term minimum. The initial retransmission interval is set to the sum of the average round trip delay and three times the difference between average round trip delay and the minimum delay. The assumption made here is that the distribution of delay times is asymmetric with the tail on the upper side being three times as long as the tail on the lower side. With this algorithm, occasional retransmissions occur. The original algorithm was of the same general form but with the three replaced with a one; a symmetric distribution was assumed. This produced fairly frequent retransmissions.

The initial retransmission rate is backed off exponentially. That is, each succeeding interval is longer by a proportional amount than the preceding one. The constant of proportionality

Any stats available?

is 30 percent. The interval is limited to a maximum of 60 seconds. The minimum interval is also limited to 500 milliseconds. Statistics are maintained on the number of retransmissions and the number of extraneous acknowledgments. If the networks do not generate duplicates, then the difference of these two is the number of packets lost. The short delay information is also available to the user.

A related but largely independent improvement was the addition of a memory verify command. This command causes the contents of the PDP-11 memory to be compared with the contents of the buffer in XNET and the differences printed. Comparison is done under the word search mask and occurs between limits supplied by the user.

The TOPS20 operating system usurps control-Q from the user's terminal to control scope page scrolling. In XNET, control-Q queries the status of network transmission and consequently users with scope terminals on TOPS20 either had to forgo control of page scrolling or the ability to query network transmission status. Control-B was added as an alternate query interrupt character to circumvent this problem.

C. Transmission Control Program (TCP)

The TENEX TCP has been revised in order to implement the most recent protocol feature. Specifically, "options", half-open connection resolution, the reset mechanism, and the user call "ABORT" are now supported.

Options provide a way of sending extended header information for which no fields have been permanently allocated. The type of information carried in options is either very seldom sent (if at all), or is used only for specific purposes and is not common to all connections. An example of the former is the "secure open" option, which is sent only in the first packet in each direction on a new connection. Timestamp and debugging label options fall into the latter category.

Half-open connections arise if the host on one end of a connection crashes. When restarted, it may attempt to reopen the same connection and the protocol must guarantee that it does not pick up the previous incarnation. Since TCP specifies that the initial sequence number carried by SYN packets is geared to time, the host which remained functional will receive a SYN which is not a duplicate of the one which established the first incarnation of the connection, but it cannot tell if it is an old, delayed duplicate of some previous incarnation. This decision must be made by a host which was restarted, and all the receiver of the questionable SYN does is to emit a normal acknowledgement. ~~for the SYN, along with a RST (reset) command.~~ Since this packet is properly sequenced, the host which restarted will find it acceptable for processing and will cancel its attempt to reestablish the connection. Further, the RST packet has enough information so that it too can generate and send an acceptable RST to the end with the half-open connection. After this is done, the previous incarnation will have been completely

this is wrong.

makes no sense.

deleted and both hosts may then go on to set up a new, consistent incarnation of the connection.

Given the mechanism described above, it became possible for a host to unilaterally abandon its end of a connection. This in effect is what happens should that host crash. The ABORT user call on the TCP provides a way for a user program to command the TCP to delete the connection. It would do this, for example, if an attempt to CLOSE a connection were taking too long.

V. HARDWARE

This quarter was not a time for major hardware work. We are awaiting delivery of the hardware to upgrade PR PDP-11 number 2, including a disk drive, which is scheduled for delivery in the months ahead. Meanwhile, occasional use of another PDP-11 system with disks will permit development and checkout of disk routines. We will also rely on cross-network debugging in the SRI testbed.

The memory to upgrade PDP-11 number 1 to 128 K core arrived and was installed and tested. This removes a serious constraint on our software development and testing efforts.

Memory was also added to each Packet Radio Digital Unit, in the amount of 1/2 K each, by Collins. This augmentation was necessary for the new CAP3 software released this quarter by Collins.

And in conclusion, a mysterious problem arose in the Very Distant Host connection between UCL's gateway and their TIP. Several areas were suspect: the RTP code driving the VDH; the VDH hardware; the modems and communication line; and the TIP itself. We supplied consultation and cooperative testing, using our gateway machine, to help UCL personnel diagnose the difficulty. Through efforts which are a credit to both organizations, the problem was isolated to a host port interface in the TIP. Appropriate TIP maintenance personnel were informed and the problem corrected.

Bolt Beranek and Newman Inc.

Report No. 3520

Command and Control Related Computer Technology
Part II. Speech Compression and Evaluation

March 1977

Submitted to:
Defense Advanced Research Project Agency

TABLE OF CONTENTS

I. INTRODUCTION

During the past quarter, we conducted informal listening tests to compare the speech quality resulting from the use of the variable frame rate (VFR) scheme for transmitting LPC data based on our automatic perceptual modeling approach, with the speech quality obtained from (1) fixed 100 frames/sec (fps) transmission, and (2) VFR scheme using our earlier log likelihood ratio method, under the following vocoder conditions: 11-pole fixed or variable-order LPC analysis, parameter quantization, VFR transmission of pitch and gain, and optimal linear interpolation. (Details of our automatic perceptual modeling scheme were given in our last Quarterly Progress Report.) Two particularly interesting results obtained from this comparative evaluation are: (1) For utterances for which LPC parameters vary relatively slowly in time, the syntheses from the 100 fps system sounded worse, in particular had a more "wobble" quality, than those produced by the perceptual modeling scheme. (2) The perceptual-model-based VFR scheme produced a better quality speech at 1700 bits/sec (bps) than did the likelihood-ratio-based scheme at 2100 bps. A detailed account of these and other results is presented in Section II.

We implemented a new class of VFR schemes for the transmission of pitch and gain. On-going informal listening tests suggest that the use of these new VFR schemes may provide a

saving of a little over 100 bps in the transmission rate over the currently used VFR schemes, without causing any perceivable change in speech quality.

Also during the last quarter, we completed the second phase of FTP (File Transfer Protocol) development, which provides the capability to reformat files and thereby allows a convenient transfer of waveform files between the PDP11 and TENEX.

In speech quality evaluation we completed the initial analysis of our factorial speech quality study described in a previous Quarterly Progress Report. In this study, four values for number of poles (13, 11, 9, 8) were combined factorially with three values of step size for quantization of log area ratios (0.5, 1, 2 dB), and with four values of frame rate (100, 67, 50, 33 fps), to define 48 LPC vocoder systems with overall bit rates ranging from 8700 down to 1300 bps. Subjects rated the DEGRADATION of signal quality by each vocoder, for each of seven sentence tokens, chosen to challenge LPC vocoders maximally. The results define the combination of LPC parameters yielding the best speech quality for any desired overall bit rate.

Also in the last quarter we issued the ARPA-NSC Note No. 97 describing the results of speech quality testing of variable frame rate LPC vocoders [8].

II. INFORMAL TESTING OF PERCEPTUAL MODELING SCHEME

It should be recalled that in our earlier developmental work on perceptual model -- both manual and automatic schemes -- we considered 14-pole fixed-order LPC analysis of unpreemphasized speech, fixed 100 fps transmission of pitch and gain, and unquantized LPC data [1]. Our conclusion based on informal listening tests was that the syntheses obtained from the manual and automatic perceptual modeling approaches and from the fixed 100 fps system had about the same overall speech quality. However, the average frame rates of log area ratio (LAR) transmission for the manual and automatic VFR schemes were only about 27 fps. An experienced listener could, for some utterances, pick the synthesis from the automatic scheme as being slightly inferior to the syntheses from the other two systems. Last quarter, considering mainly the automatic perceptual modeling scheme, we repeated the comparative informal quality tests under vocoder conditions similar to those of ARPA-LPC System II [2]. In an attempt to understand the effect on speech quality of each of these vocoder conditions, we considered them one by one and tested the resulting vocoder systems. All the syntheses included in the informal listening tests were generated using our previously reported all-pass excitation method [3]. As speech material, we used the 9 utterances: AR4, DD2, DD6, DK4, DK6, JB1, JB5, RS3, and RS6, from three male (DD, DK and JB) and two female (AR and RS) speakers.

In the discussions that follow, by perceptual modeling scheme we mean the automatic scheme.

A. Comparison with Fixed-Rate Transmission Scheme

Test results reported in this section involve the vocoder systems 1 through 5, listed in Table 1.

For Vocoders 1 and 2, we used unquantized LPC parameter data, fixed 12-pole LPC analysis of unpreemphasized speech, and fixed 100 fps transmission of pitch and gain. Vocoder 1 transmitted LARs at a fixed rate of 100 fps, while Vocoder 2 used the VFR scheme based on our perceptual model to yield an average frame rate of about 27 fps. The two vocoders were found to have about the same overall speech quality, with Vocoder 1 producing a slightly better clarity than Vocoder 2 (specific examples: the words "trouble" and "drown" in DK6).

Vocoders 3 and 4 given in Table 1 considered parameter quantization (pitch: 6 bits, gain: 5 bits, and LARs: 44 bits for voiced sounds and 42 bits for unvoiced sounds), fixed 11-pole analysis of preemphasized speech, and fixed 100 fps transmission of pitch and gain. Fixed-rate Vocoder 3 produced a bit rate of 5650 bps, while the variable-rate Vocoder 4 produced an average bit rate of about 2450 bps. Through informal listening tests on the syntheses from these two vocoders, we observed two interesting aspects. First, the word "drown" in DK6 synthesized

Vocoder #	Quanti-zation	LPC Order	Transmission		VFR Scheme for LARs	Average Bit Rate
			Pitch and Gain	LARs		
1	No	Fixed	Fixed	Fixed	-	-
2	No	Fixed	Fixed	Variable	PM	-
3	Yes	Fixed	Fixed	Fixed	-	5650
4	Yes	Fixed	Fixed	Variable	PM	2450
5	Yes	Variable	Variable	Variable	PM	1700
6	Yes	Variable	Variable	Variable	LLR	2100

Table 1. Fixed-rate and variable-rate vocoders included in the comparative evaluation study (PM: perceptual model, LLR: log likelihood ratio method).

from Vocoder 4 was perceived more like "drawn", unlike the syntheses from the first three vocoders, all of which sounded clearly as "drown". In view of the differences between Vocoders 1 through 4 (see Table 1), we conclude that the above distortion was the result of not just VFR transmission of LARs, or just parameter quantization, but a combination of the two factors. The interaction between the two factors produces, for the untransmitted data frames, interpolated LAR values that are in general different from those obtained using unquantized LARs. The above incident represents one example of cases where this difference in interpolated LAR values leads to a perceivable effect. The second interesting result of our informal listening tests is that, for the slowly varying utterances JB1 and DD2, the syntheses from the 5650 bps fixed-rate Vocoder 3 actually sounded worse, in particular had a more "wobble" quality than from the 2450 bps VFR System 4. We had had the same experience with our earlier VFR scheme that uses the log likelihood ratio criterion. Our explanation for the observed quality difference is that for slowly varying utterances, the error due to parameter quantization is more than the error due to parameter interpolation. The noted quality difference was quite perceivable when listening through earphones and somewhat diminished when listening through loudspeakers.

We added the optimal linear interpolation feature [4] to Vocoder 4 in an attempt to improve the clarity of vocoded speech

in places like "trouble" and "drown" in DK6. The interpolation scheme increased the bit rate by about 50 bps, but did not produce any perceivable quality change.

Vocoder 5 given in Table 1 differs from Vocoder 4 in two respects: variable-order LPC analysis (maximum order = 11 poles) was used, and pitch and gain were transmitted at a variable frame rate [5]. The average transmission rate was reduced from about 2450 bps for Vocoder 4 to about 1700 bps for Vocoder 5, due to those two changes. The quality differences between the two vocoders were quite small, with speech from Vocoder 5 sounding a bit "crispier" than speech from Vocoder 4.

B. Comparison with Log Likelihood Ratio Method

Variable-rate Vocoders 5 and 6 given in Table 1 have the same features except for the VFR scheme used for LAR transmission: Vocoder 5 employed the perceptual modeling scheme, while Vocoder 6 used our earlier log likelihood ratio method. The average frame rates of LAR transmission, obtained for the two vocoders, for each of the 9 utterances are given in Table 2. The average frame rate over the 9 utterances was found to be about 27 fps for the perceptual modeling scheme and about 36 fps for the likelihood ratio method. The average bit rates for the two vocoders were, respectively, about 1700 bps and 2100 bps. Informal listening tests, however, showed that Vocoder 5 with the perceptual-model-based VFR scheme actually produced a <u>better</u>

Utterance	Frame Rate for LARs	
	PM	LLR
AR4	35.4	44.3
DD2	24.7	27.6
DD6	30.8	36.7
DK4	24.3	36.0
DK6	23.4	33.6
JB1	18.1	31.9
JB5	31.5	43.0
RS3	27.6	30.1
RS6	23.4	37.9
Average	26.6	35.7

Table 2. Average transmission frame rates of LARs for 9 utterances, obtained using the two types of VFR schemes: 1) Perceptual model (PM), and 2) Log likelihood ratio (LLR) method.

quality speech than did Vocoder 6 with the likelihood-ratio-based VFR scheme. Specific problems in the syntheses from Vocoder 6 were: "troublith" instead of "trouble with" in DK6, "around'n" instead of "around on" in JB5, a "pop" sound during "trouble" in DD6, and slightly more "wobble" quality in DD2 and JB1.

The above results clearly suggest the superiority of the perceptual modeling scheme over the likelihood ratio method. We plan to run a formal subjective quality test to confirm these results.

III. FIT: A NEW TRANSMISSION SCHEME FOR PITCH AND GAIN

In the past we have used single-threshold and double-threshold VFR schemes for the transmission of pitch and gain [5]. (As LPC gain parameter, we transmit per-sample energy in decibels of the unpreemphasized speech.) The single-threshold scheme transmits the parameter value (pitch or gain) for a given frame if the absolute difference between the value and the preceding transmitted value exceeds a prespecified threshold. The double-threshold scheme follows the same rule, except that it will instead transmit the parameter value for the frame immediately preceding the present frame if the above absolute difference exceeds a prespecified second (higher) threshold; this avoids the need to do parameter interpolation at the receiver between largely different data frames. We previously recommended the use of specific double-threshold VFR schemes on quantized pitch and gain data for ARPA-LPC System II [5]. These schemes would reduce the average transmission frame rate from the analysis rate of 100 fps to about 35 fps for pitch and 32 fps for gain.

The above-mentioned single-threshold scheme is similar to the so-called "floating-aperture predictor" (or FAP, for our discussion) which has been used for data compression in telemetry applications [6,7]. The main difference between the two is in the way data reconstruction takes place at the receiver i.e., how

the untransmitted parameter values are approximated. FAP employs a stair-step reconstruction in that a transmitted value is held constant for all the frames up to the next transmission, where the value is instantaneously updated to be the next-transmitted value. Our single-threshold scheme, however, performs linear interpolation between adjacent transmitted values to generate a smoother approximation. (The double-threshold scheme has the same feature, except that, as mentioned above, it produces less interpolation error at the expense of a slight increase in frame rate.) It is felt that in speech resynthesis applications the smooth approximation produced by interpolation should produce less speech quality distortion (e.g., "roughness") than the stair-step approximation used in the FAP method. However, at the transmitter, our VFR scheme (hereafter loosely called as FAP scheme) does not explicitly take advantage of the fact that the receiver performs linear interpolation for data reconstruction. The inclusion of this feature may perhaps yield further data compression. To this end, we have adapted the so-called "fan interpolation" technique (abbreviated here as FIT) that has been used once again in telemetry applications [6,7].

A. Single-Threshold Scheme

The FIT method previously used in the literature [6,7] is indeed a single-threshold scheme. The method relies on the approximation of the analysis or source data by straight line

segments and transmits only those parameter values corresponding to the end frames of these segments. Given some initial transmitted frame, it finds the longest line for which the maximum error magnitude between the line and the data over the length of the line is below a given threshold. We treated the case where quantized parameter values (levels) are used for deciding when to transmit. In computing the error between the quantized parameter level for a frame and the interpolation line, we compute the interpolated value for that frame, round it off to the nearest (integer) level and then find the difference between this and the actual quantized parameter level for that frame. (Rounding is done such that if the fractional part of the interpolated value is equal to or greater than 0.5 then it is rounded up, otherwise it is rounded down.) As before, untransmitted frames are indicated by transmitting a zero header bit. At the receiver, quantized levels for untransmitted frames are generated by interpolating between the adjacent transmitted levels and rounding off the interpolated value to the nearest level as explained above.

A step-by-step description of the FIT single-threshold scheme is given below, where I_n denotes the quantized level of the parameter for frame n, the symbol [ ] refers to the above rounding operation, and T is the preselected threshold.

(1) Transmit value at frame n

$m \leftarrow 2$

(2) $k \leftarrow 1$

(3) $P \leftarrow (m-k)/m \; I_n + k/m \; I_{n+m}$

$E \leftarrow |[P] - I_{n+k}|$

If $E \leq T$, go to (4)

$n \leftarrow n+m-1$

Go to (1)

(4) $k \leftarrow k+1$

If $k \leq m-1$, go to (3)

(5) (No transmission)

$m \leftarrow m+1$

Go to (2)

It is clear from step (3) that with frames n and (n+m) as end frames, the scheme looks at the magnitude of the interpolation error, in order, from frame (n+1) to (n+m-1) and decides to transmit frame (n+m-1) value at the first instance the error magnitude exceeds T.

If T=0, it is easily seen that the receiver has the same parameter data as at the output of the quantizer. As mentioned earlier, the same result is also achieved using the FAP method with a zero threshold and with stair-step reconstruction at the

receiver. Average transmission frame rates produced by the two methods can, however, be different; the extent of this difference depends upon the nature of the data, in this case quantized parameter levels. For instance, if the data has frequent occurrence of sequences of equal levels (i.e., presence of horizontal or level lines), then the FAP scheme would generally do better yielding a lower frame rate than the FIT method; the reason for this is that the latter method transmits both end frames for each level line, while the former transmits only the first end frame. On the other hand, if the data involves a large number of sloped or nonlevel lines then the opposite result is true in that the FIT method yields a lower frame rate.

Experimental results obtained using the above FIT method on quantized pitch and gain are reported in Subsection C below.

B. Double-Threshold Scheme

The double-threshold version of the FIT method operates as follows. Assume that frames n and (n+m) are the end frames of the interpolation line under consideration. Then, (1) if the maximum interpolation error magnitude over the length of the line exceeds the second (higher) threshold T2, then frame (n+m-1) value is transmitted; (2) if the maximum error magnitude exceeds the first (lower) threshold T1, and not T2, then frame (n+m) value is transmitted; (3) if the maximum error magnitude does not exceed T1, then a new interpolation line is considered between

frames n and (n+m+1), and the entire procedure is repeated. A step-by-step description of the double-threshold scheme is given in the next page.

For our earlier VFR scheme (FAP), the motivation to use the double-threshold scheme has been to improve the accuracy of parameter interpolation performed at the receiver between adjacent transmitted values. The same motivation does not hold for the above FIT method, since it explicitly considers interpolation error as part of its transmission strategy. Why, then, should one consider the FIT double-threshold scheme? The answer may be given as follows. Considering quantized parameter data, the FIT single-threshold scheme allows only integer thresholds. In effect, the double-threshold scheme may be viewed as equivalent to a single-threshold scheme that can allow a noninteger threshold. For example, the (0,1) double-threshold scheme produces average frame rate and speech quality that lie between those of the two single-threshold schemes with thresholds 0 and 1. This point will be more clear from the experimental results provided in the next subsection.

(1) Transmit value at frame n

$m \leftarrow 2$

(2) $Flag \leftarrow 0$

$k \leftarrow 1$

(3) $P \leftarrow (m-k)/m\ I_n + k/m\ I_{n+m}$

$E \leftarrow |[P] - I_{n+k}|$

If $E \leq T2$, go to (4)

$n \leftarrow n+m-1$

Go to (1)

(4) If $E \leq T1$, go to (5)

$Flag \leftarrow 1$

(5) $k \leftarrow k+1$

If $k \leq m-1$, go to (3)

(6) If $Flag = 0$, go to (7)

$n \leftarrow n+m$

Go to (1)

(7) (No transmission)

$m \leftarrow m+1$

Go to (2)

<u>Description of our FIT double-threshold scheme</u>

C. Experimental Results

Below, we report experimental results obtained using the FIT method on the quantized pitch and gain data. Our speech data base consisted of a total of 11 utterances, representing about 25 seconds of speech, from 5 male and 5 female speakers. This data base is the same as the one used for computing average transmission frame rate data for our earlier FAP-type VFR schemes [5].

Pitch:

The FIT single-threshold scheme produced average frame rates of 35, 18 and 14 fps for values of the threshold T=0, 1 and 2, respectively. Using the (0,1) double-threshold scheme, we obtained an average frame rate of 26 fps. This latter rate should be compared against the rate of 35 fps that we had reported for our earlier (0,1) FAP scheme [5].

Gain:

The FIT single-threshold scheme produced average frame rates of 57, 31 and 22 fps for values of the threshold T=0, 1 and 2, respectively. Using the FIT double-threshold scheme, we obtained average frame rates of 26 and 19 fps for the two thresholds (T1,T2)=(1,2) and (2,3), respectively. In contrast, the (2,3) double-threshold FAP scheme produced an average frame rate of 32 fps [5].

Informal listening tests are under way to compare the FAP and FIT schemes for pitch and gain transmission in terms of the quality of vocoded speech. The use of the (0,1) FIT scheme for pitch transmission and the (2,3) FIT scheme for gain transmission in ARPA-LPC System II, instead of the corresponding FAP schemes currently being used, would mean a total saving of about 119 bps in the transmission bit rate, 54 bps (=6 bits x 9 fps) from pitch transmission and 65 bps (=5 bits x 13 fps) from gain transmission. Also, the average transmission frame rates associated with these new pitch and gain transmission schemes (26 and 19 fps, respectively) are in line with the minimum necessary frame rate of LAR transmission of about 25 fps or about 2 transmissions per phoneme that we reported in our perceptual model work [1].

IV. REAL-TIME IMPLEMENTATION

We completed the second phase of FTP development, which provides the capability to reformat files and thereby allows a convenient transfer of waveform files between the PDP11 and TENEX. Handlers and utility programs for using the IMLAC as a peripheral to the PDP11 have been finished. A set of programs to be used for backing up the PDP11 system on TENEX have also been implemented. The initial A/D spooling program is currently being modified to support waveform display and editing functions on the IMLAC. We plan to implement a playback program on the PDP11, which would allow us to conveniently and rapidly prepare audio tapes containing specified sequences of synthesized and natural speech utterances for demo purposes and for subjective speech quality tests.

V. SUBJECTIVE QUALITY EVALUATION

A. Introduction

Our factorial subjective-quality study was performed to measure how the quality of LPC vocoded speech is affected by three different methods of reducing bit rate. These were:

1) reducing the number of poles used for spectral matching,
2) coarsening the step size used in quantizing the coefficients (log area ratios, see [9])
3) reducing the number of frames of coefficients transmitted per second.

The procedures we followed are summarized here, for convenience. To establish the best operating point, for a range of different bit rates, it was necessary to perform a factorial study, in which each value of a parameter occurred with every combination of values of the other parameters. We used the following set of parameter values: Number of Poles, P: 13, 11, 9, or 8; Quantization Step Size, Q: 0.5, 1.0, or 2.0 dB; and Frame Rate, R: 100, 67, 50, or 33 per second, yielding 48 LPC svstems (4x3x4). Two additional systems were included. One was an LPC system with 13 poles, quantization step size of 0.25 dB, and transmission rate of 100 frames per second. The other consisted of PCM speech at 110 kbps (i.e. the waveform sampled at 10 kHz and quantized to 11 bits), to act as an undegraded anchor.

The bits per frame for each combination of number of poles and quantization step size appear in Table 3.

Pitch and gain were transmitted at the same frame rate as the coefficients. The expected overall bit rate for any system is calculated by adding 6 bits of pitch coding and 5 bits of gain to the bits per frame, and multiplying by the appropriate frame rate. The measured overall bit rate of the LPC systems ranged from 8430 bps (P = 13, Q = 0.25 dB, R = 100/sec, expected rate = 8700 bps), down to 1225 bps (P = 8, Q = 2.0 dB, R = 33/sec, expected rate = 1267 bps). Note that these rates do not include the benefits of Huffman coding, in which the most frequently used values are assigned the shortest codes. This procedure could further reduce bit rates by about 20%, with absolutely no change in the coefficient values transmitted [10].

B. Sentence Materials

Our earlier subjective quality tests showed the necessity of passing all sentence materials through all systems [11]. Other researchers have reached similar conclusions [12]. In our earlier tests, we developed a set of six sentences, each read by six talkers, that was both representative, in that it covered a wide range of speech events and talker characteristics, and also challenging, in that some speech material was included that would fully extend any LPC vocoder's abilities. Unfortunately, we

Quantization Step Size	No. of Poles 13	11	9	8
0.25 dB	76	--	--	--
0.5 dB	63	55	47	43
1.0 dB	50	44	38	35
2.0 dB	37	33	29	27

Table 3: Expected bits per frame for all combinations of number of poles and quantization step size used in the present study (excluding pitch and gain).

ID	FO	Sentence
JB1	119	Why were you away a year, Roy?
DD2	134	Nanny may know my meaning.
RS3	195	His vicious father has seizures.
AR4	165	Which tea-party did Baker go to?
JB5	124	The little blankets lay around on the floor.
DK6	97	The trouble with swimming is that you can drown.
RS6	193	The trouble with swimming is that you can drown.

Table 4: The seven stimulus sentences, with the speaker's average fundamental frequency in Hz.

could not use all 36 speaker-sentence combinations in the present study, since passing them through all 50 vocoder systems would have made the study unmanageably large. We therefore selected a subset of seven speaker-sentence combinations, and confirmed that they were adequately representative of the full set by repeating the MDPREF analysis using just the data from the subset. The multidimensional solution obtained from the subset was substantially the same as that obtained from the complete set.

The subset of sentence tokens that was selected consisted of: JB1, DD2, RS3, AR4, JB5, DK6, and RS6, where the initials identify the speaker and the number identifies the sentence. Relevant details of the sentences, and of the speakers' voices, are given in Table 4.

C. Generation of Stimulus Tapes

Each of the seven input sentences was digitized (11 bits, 10 kHz), and passed through each of the 50 simulated vocoder systems, to yield a total of 350 different stimulus items.

Earlier studies have demonstrated that a subject's judgment, especially of speech stimuli, can be strongly affected by the preceding stimulus (e.g. [13]). It is important to control for effects such as this by counterbalancing the presentation order. A complete counterbalancing of the 50 vocoder systems was

generated, in which every system followed every other system once, with independent approximate counterbalancing of the sentences. This required only seven passes through the 350 stimuli, and had the further advantage that even within each pass, all ranges of contrast between successive systems occurred equally often, so that no severe departures from balance occurred even within one pass. The sequence was generated by a trial and error search, following an algorithm described by Williams [14]. No system and no sentence followed itself.

We tried to further reduce sequence effects, and thus improve the reliability of the data, by the method described in QPR #8. A continuous speech babble, at the same level as the speech, was automatically faded in and out again during the inter-stimulus interval. We hoped that, by analogy with the "suffix" effect found in studies of auditory short term memory [15], the babble would interfere with the memory trace of earlier stimuli, on which sequence effects presumably depend. The babble was developed at BBN for other purposes [16]. The babble signal was recorded on a separate track of the tape, to permit the signal to be played with or without the babble.

Seven experimental tapes were then recorded. Stimuli were presented in blocks of ten, at a rate of one every 7.5 seconds, with a longer gap between blocks.

D. Experimental Procedures

The subject's task was to rate the degradation of the stimuli he heard. This negative attribute was chosen for scaling, as in our earlier experiments, because the scale has a natural origin, or zero, corresponding to undegraded speech. Instead of assigning a number to his judgment, the subject made his response by making a mark on a 10 cm line on his answer sheet. Two visual anchors were provided on the response line. The left anchor was 4 mm from the left end of the line, and was marked "PERFECT". The right anchor was 1 cm from the right end of the line. For data analysis, the response was converted into the distance in millimeters from the left end of the line (not the anchor) to the subject's mark where it crossed the response line. Thus the degradation ratings range between 0 and 100, with small numbers corresponding to high quality, and large numbers to poor quality.

Nine subjects served in the experiment. They were recruited by local university summer placement offices: all reported having normal hearing. All of the subjects made the first two passes through the 350 stimuli, and three of them made a further three passes each. The first 200 stimuli (20 blocks of 10) were repeated exactly, after the first two passes through the 350 stimuli, without the subjects knowledge. The data generated by this duplication, and an equivalent one performed after pass 5,

were used to provide a check on how reliably a subject could assign ratings, and on the stability of his judgements.

E. Results

First, to check on the reliability of the data, the responses collected on each pair of passes through the 350 stimuli were correlated, for each subject. The correlation coefficients are shown in Table 5. From scatter plots of the first vs. the second pass, for each subject, it became apparent that some of the high correlations might be inflated by the fact that the undegraded anchor stimuli (System 000) were rarely confused with any of the other systems, but received very low degradation ratings (see below in Table 6). Therefore, the correlations between pairs of passes were recalculated, excluding the anchors. These correlations coefficients appear in the right hand column of Table 5. With an N of 350, a correlation coefficient of 0.4 is significant at P<.001, and one of 0.6 is significant at P<.000,000,01 !

As can be seen from Table 5, all correlations were significant well beyond P<.001. Therefore, although there was some variability between subjects, all the subjects apparently gave highly reliable data.

Subject 1	Pass 1 vs 2	.827	.784
Subject 1	Pass 1 vs 3	.819	.769
Subject 1	Pass 1 vs 4	.834	.788
Subject 1	Pass 1 vs 5	.794	.735
Subject 2	Pass 1 vs 2	.484	.446
Subject 3	Pass 1 vs 2	.713	.677
Subject 4	Pass 1 vs 2	.535	.379
Subject 5	Pass 1 vs 2	.619	.576
Subject 5	Pass 1 vs 3	.633	.601
Subject 5	Pass 1 vs 4	.633	.606
Subject 5	Pass 1 vs 5	.642	.613
Subject 6	Pass 1 vs 2	.539	.403
Subject 6	Pass 1 vs 3	.509	.347
Subject 6	Pass 1 vs 4	.563	.425
Subject 6	Pass 1 vs 5	.542	.365
Subject 7	Pass 1 vs 2	.628	.613
Subject 8	Pass 1 vs 2	.758	.663
Subject 9	Pass 1 vs 2	.782	.773

Table 5: Correlations between first and other passes, including (left) and excluding (right) the undegraded anchors.

The mean degradation rating was calculated for each system, both by sentence, and pooled across all seven sentences. These mean degradation ratings are shown in Table 6, and are plotted in Figures 1 to 6. Each system is identified by three digits, corresponding to the parameter level for P, Q, and R, respectively. Thus system 231 used level 2 of P (11 poles), level 3 of Q (1.0 dB) and level 1 of R (100/sec), as shown in the key to the figure. System 000 corresponds to the 110 kbps PCM speech, used as undegraded anchor. The mean ratings (N.B. not the ratings) have standard deviations ranging between 1.0 and 1.7 degradation points. Therefore any difference between two plotted means that is larger than about 4-5 points is probably significant at $P<.0.05$.

A more sensitive test of the difference between two systems is possible. For any pair of systems, a pair of ratings is available for each of the 7 sentences, for each of the 27 passes through the 350 stimuli. If the pair of systems yielded equally degraded speech, the differences between the members of each pair of ratings should have zero mean. A t-test based on this logic was performed, and its results are presented in cryptic form in Table 7. The table consists of a 50x50 matrix. The left column and the top three rows contain System-IDs for indexing the rows and columns of the matrix, respectively. The column indexes should be read vertically. The cell entries in the table are the

PQR	JB-1	DD-2	RS-3	AR-4	JB-5	DK-6	RS-6	POOLED
000	14.4	16.3	17.1	6.4	7.9	11.5	11.9	12.22
111	44.2	44.3	52.9	51.6	28.1	34.9	46.4	43.22
121	54.0	49.8	58.3	52.0	30.6	34.9	51.3	47.27
122	51.5	42.7	66.6	58.3	33.7	32.7	53.4	48.42
123	45.6	50.9	72.0	60.1	31.9	34.1	53.6	49.75
124	50.2	52.1	72.8	67.7	52.9	45.7	62.3	57.68
131	60.6	45.9	59.6	54.7	37.9	34.7	62.7	50.86
132	57.7	53.2	63.4	59.8	38.7	30.0	55.1	51.12
133	51.1	53.6	70.9	58.7	34.5	32.9	57.7	51.34
134	51.1	52.4	73.0	70.9	57.8	45.8	68.8	59.96
141	70.3	63.5	68.4	62.5	56.2	53.7	66.6	63.04
142	71.0	63.5	70.5	61.2	55.6	59.3	59.7	62.97
143	71.8	63.2	72.8	60.0	47.1	43.8	64.6	60.46
144	59.9	64.1	75.1	70.6	52.4	44.2	66.0	61.78
221	56.3	50.4	54.9	50.6	29.6	37.7	54.0	47.65
222	51.8	48.3	67.2	57.6	39.9	30.2	52.1	49.58
223	55.0	52.4	68.9	63.1	38.6	34.9	59.4	53.19
224	49.0	55.1	73.4	65.0	48.4	41.0	66.0	56.86
231	60.5	48.9	60.9	53.8	35.4	32.5	55.2	49.60
232	52.2	53.9	62.0	56.6	43.7	42.5	53.0	51.97
233	51.4	49.2	72.1	62.3	41.7	48.2	55.0	54.27
234	53.5	56.6	72.0	69.8	50.1	41.8	63.7	58.21
241	71.7	63.9	62.4	59.4	49.2	48.6	63.0	59.75
242	69.9	59.7	71.9	62.9	49.9	53.2	59.4	60.99
243	68.2	58.0	69.3	61.7	44.4	42.1	63.1	58.14
244	67.5	67.9	74.4	69.2	60.1	44.3	70.9	64.89
321	66.8	58.8	58.5	53.7	46.4	57.0	52.5	56.24
322	68.4	53.9	68.4	62.3	59.1	57.6	50.3	60.01
323	67.0	57.0	74.1	61.1	52.6	64.4	56.5	61.82
324	70.3	64.6	75.0	70.9	66.4	69.9	65.7	68.95
331	72.8	61.4	59.5	57.0	51.1	57.9	58.5	59.75
332	61.7	59.6	66.4	60.6	52.9	61.5	54.7	59.63
333	74.5	62.2	69.4	59.0	56.5	59.7	61.2	63.22
334	69.9	68.8	76.2	68.9	69.6	69.6	73.9	70.97
341	76.1	73.4	67.6	60.7	57.2	63.6	60.3	65.56
342	75.4	72.1	70.0	67.8	56.6	64.0	61.1	66.72
343	72.1	74.7	72.7	69.9	57.0	63.4	63.5	67.62
344	71.4	75.3	74.3	68.1	71.6	64.4	70.6	70.83
421	79.0	59.9	56.9	56.7	63.9	76.2	54.6	63.86
422	80.4	68.7	64.4	62.7	66.6	75.7	55.7	67.76
423	79.3	65.9	71.8	63.4	68.4	74.4	62.7	69.41
424	81.6	69.9	70.0	71.7	74.7	76.9	66.9	73.07
431	77.9	63.5	61.1	56.2	63.9	69.4	59.4	64.48
432	76.6	67.0	68.3	63.8	66.0	78.0	53.0	67.52
433	76.0	61.7	69.9	62.7	65.6	76.9	59.0	67.38
434	80.0	72.9	76.2	70.7	75.6	77.4	71.7	74.92
441	81.4	64.0	69.2	66.9	67.0	75.6	64.7	69.85
442	80.4	72.5	71.7	68.9	65.6	77.9	60.7	71.10
443	78.2	66.9	74.0	68.9	68.6	77.8	71.0	72.19
444	78.0	71.1	76.9	71.9	79.5	82.6	69.4	75.63

Table 6: Mean Degradation Rating, by Sentence

```
 P   01111111111111222222222222333333333333444444444444
  Q  01222233334444222233334444222233334444222233334444
   R 01123412341234123412341234123412341234123412341234

000  \7777777777777777777777777777777777777777777777777
111  3\877777777777777777777777777777777777777777777777
121  32\--799877777--77-8777777777777777777777777777777
122  33-\-799977777--77-8777777777777777777777777777777
123  33--\7---77777--87--777777777777777777777777777777
124  33333\333-7797332-331--9-7--87--777777777777777777
131  3311-7\--777771--7--97777777777777777777777777777
132  3311-7-\-777771--7--97777777777777777777777777777
133  3321-7--\777771--7--87777777777777777777777777777
134  33333-333\89--3331333----7---7--97777797778777777
141  3333333332\-2-3333333213-32-722-7-887-777-88777777
142  3333333331-\1-3333333213-32-722-79877-777-77777777
143  333331333-89\-3332333---172--7--977777977787777777
144  333333333----\33333332--293--7---78777-777-7777777
221  33---799977777\-77-8777777777777777777777777777777
222  33---7---77777-\87--777777777777777777777777777777
223  333328---7777732\82--77779777777777777777777777777
224  33333-33397787332\331-98-7-97799777777777777777777
231  33---7---77777--87\-777777777777777777777777777777
232  3322-7---777772--7-\977777877777777777777777777777
233  3333391127777733-931\87787-77777777777777777777777
234  33333-333-77-8333-332\-9-7--87--777778777777777777
241  33333-333-88--3331333-\--72--7--977777977777777777
242  333331333-99--33323331-\282--7---77777977787777777
243  33333-333-7798333-332--8\7--87--777777777777777777
244  3333333333--3133333333323\331823-7--97-987-9-77777
321  33333-333-7787331-32--88-7\97789777777777777777777
322  33333-333-88--3331333----71\-7--87777787777777777
323  333332333-----33333332--293-\7---78777-77797777777
324  33333333333333333333333332333\33391---2--72--7-987
331  33333-333-88--3331333----82--7\-977777877777777777
332  33333-333-88--3331333----71--7-\877777877777777777
333  3333333331--1-33333331-3-32-712\79777-777-7777777
334  333333333333333333333333333331333\332-31-93127---7
341  3333333333-13233333333333-33293317\--7-987-9-77777
342  333333333322333333333333333-333-3337-\-8--97---78777
343  333333333323333333333333331333-3338--\81--71--79777
344  33333333333333333333333333333-333-322\3--93117---7
421  3333333331--1-33333332113-32-822-7--97\777-7877777
422  33333333333333333333333331333-33391---3\-72--7-877
423  33333333333333333333333332333-333-21--3-\73--7-987
424  3333333333333333333333333333333331333133 3\333-21-9
431  3333333332--2-33333333323-331833-7--97-877\8977777
432  3333333333233333333333333331333-33391--93--72\-79777
433  33333333332333333333333333-333-3338---92--71-\78777
434  33333333333333333333333333333333333333333-333\332-
441  33333333333333333333333333333-333-321-3--83127\--7
442  333333333333333333333333333331333-333-32193337-\-7
443  333333333333333333333333333332333-333-332-3338--\7
444  33333333333333333333333333333333333333333331333-333\
```

Table 7: Results of t-tests comparing all pairs of systems (see text for more detail). A cell entry of 1,2,3 indicates that the system indexing the column yields better speech quality (less degradation) than the system indexing the row. Cell entries of 9,8,7 indicate that row quality is better than column quality. A dash indicates that the quality difference is not significant; a 1 or 9 shows the difference is significant at P<0.05; a 2 or 8 at P<0.01; a 3 or 7 at P<0.001.

digits 1,2,3, and 9,8,7. The digits 1 and 9 indicate that the two systems indexing the entry yield different speech quality, at $P<0.05$. Similarly, the digits 2 and 8, and 3 and 7, indicate differences at $P<0.01$ and $P<0.001$ respectively. A dash (-) indicates that the difference was not significant. The digits 1,2, and 3 indicate that the system indexing the Column of the table yields higher quality (less degradation) than the system indexing the Row. To take an example: the second row of the table, indexed by system 111, contains an 8 in the third column (to the right of the '\'), indexed by system 121. Since the digit is not a 1,2, or 3, the row system (111) has better quality than the column system (121), and the difference is significant at $P<.01$. Table 7 can be used in this way to determine whether any pair of systems plotted in the figures yield significantly different quality.

In Figure 1, a line joins the "best" systems using 13 poles, and other lines join the best systems using 11, 9, and 8 poles. From inspection of Figure 1, it is clear that 13-pole systems give (slightly) better quality than 11-pole systems for most bit rates above 2750. 11-pole systems are (slightly) superior between about 1500 bps and 2750 bps. These differences are small, however, and are probably not significant. The best 11 and 13 pole systems are substantially better than the best 8 or 9 pole systems at comparable bit rates. These differences are

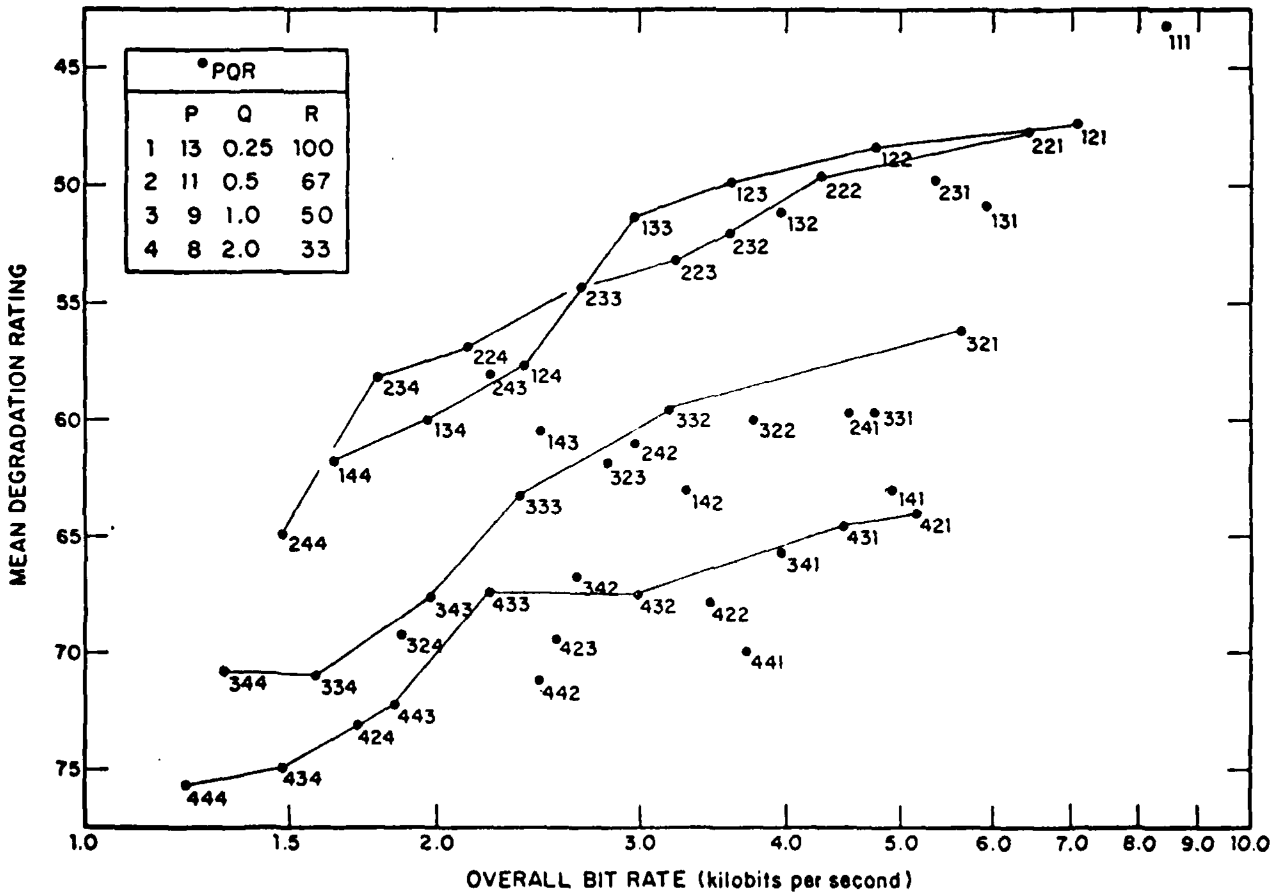

Figure 1: Mean degradation rating vs. Bit Rate for 48 LPC vocoders. Lines join "best" systems for each No of Poles.

large and highly reliable. The reason is that there is a highly significant interaction between the sex of the talker (or the talker's fundamental frequency) and the number of poles. This confirms earlier findings (Huggins & Nickerson, 1975; Huggins et al, 1976). Averaging ratings across all systems with the same number of poles shows that reducing the number of poles from 13 to 8 had relatively little effect on quality, for the three sentences spoken by females (RS3, AR4, RS6), whereas there is a massive reduction of quality for male voices when the number of poles is reduced below 11.

Figures 2 and 3 present comparable plots, with best systems joined for each level of quantization, and for each level of frame rate, respectively. The differences in quality between different levels of quantization, at a given bit rate, are significant only at the very low bit rates. Here, quality is less affected by coarsening quantization than by using fewer poles.

Figure 3 shows that below 4.5 kbps, quality can be substantially improved, at no extra cost in bit rate, by reducing the frame rate and increasing the number of bits per frame, that is, by improving "static" spectral accuracy at the expense of "dynamic" spectral accuracy. Most of these quality differences, due to changing frame rate without changing overall rate, are highly significant. The size of the effect of frame rate lends

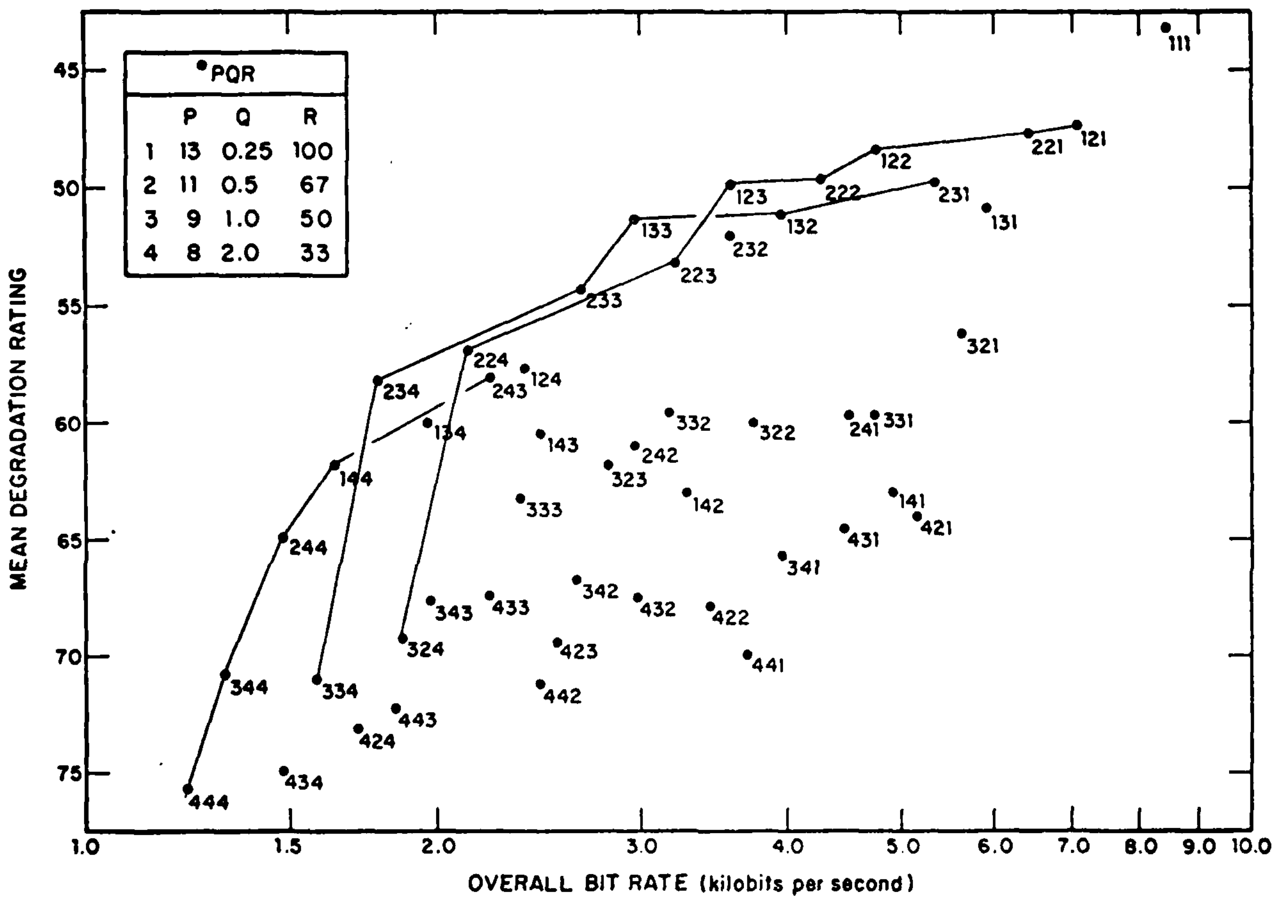

Figure 2: Degradation vs. Bit Rate. Lines join "best" systems for each Quantization Step Size.

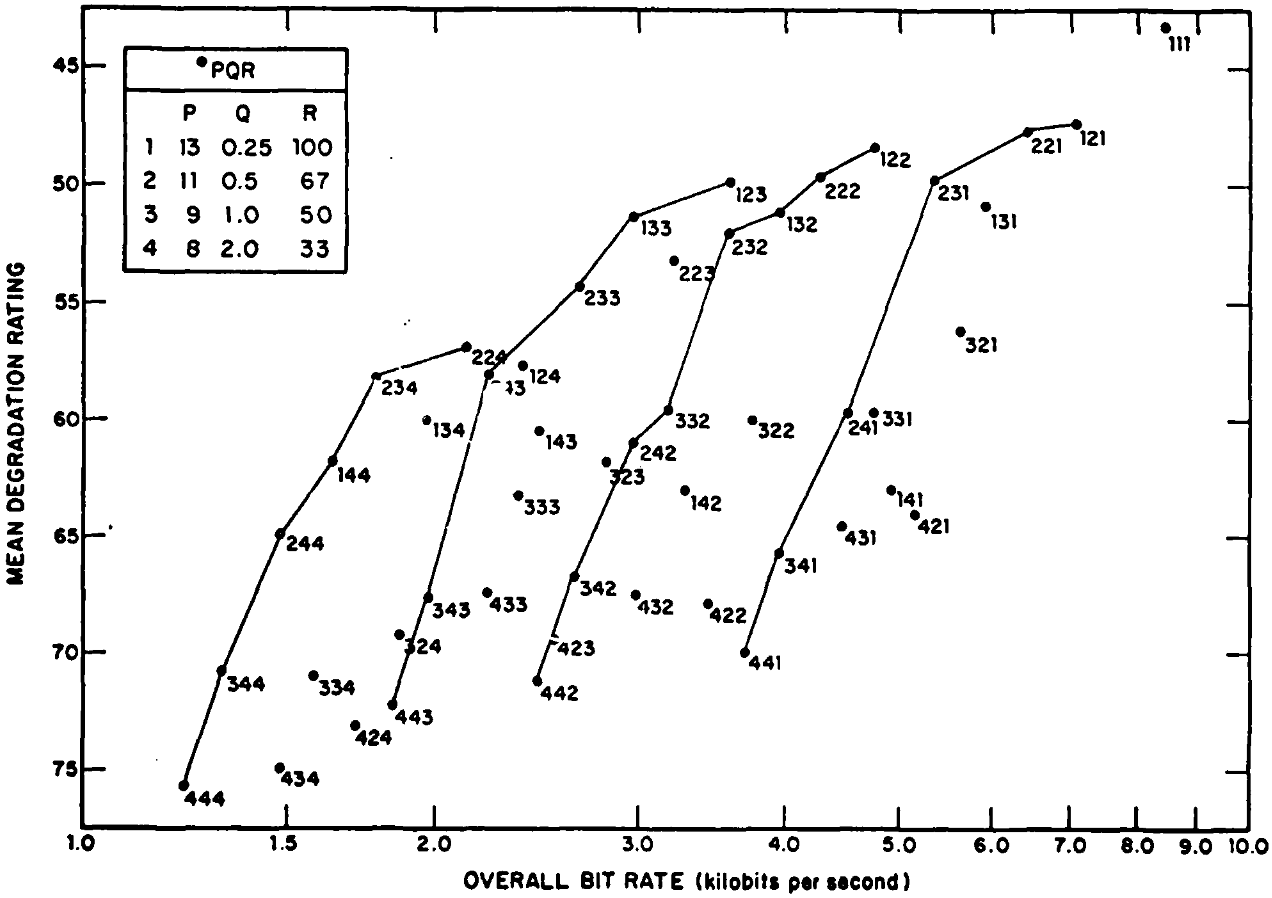

Figure 3: Degradation vs. Bit Rate. Lines join "best" systems for each Frame Rate.

further support to our earlier result (Huggins et al, 1976), suggesting that a well designed variable frame rate transmission scheme should yield substantial savings in bit rate without appreciable loss of quality.

Figures 4 to 6 lead to similar conclusions. Here, the same data points are plotted as in Figures 1 to 3, but the points are connected differently. In Figures 4 to 6, lines are drawn to show the effects on degradation of decreasing bit rate by: a) reducing the number of poles (Fig. 4), or b) coarsening the quantization step size (Fig. 5), or c) decreasing the frame rate (Fig. 6). In each case, the two remaning parameters are held constant. Comparing the slopes of the lines in the three figures shows dramatically that reducing the frame rate (Fig. 6) yields the largest savings of bit rate for the smallest loss of quality, and that for many of the systems the loss of quality shows no knee, even at the lowest frame rate. A reasonable interpretation of this is that even lower frame rates might yield acceptable quality under some conditions -- which is exactly the thrust of our work with variable-rate systems.

Secondly, inspection of Figure 4 shows that the rate of quality loss per bit saved is largest for savings gained by reducing the number of poles. In this case there is a sharp knee in most of the functions at 11 poles -- it is unfortunate that we did not also include 10 poles, although our other work suggests

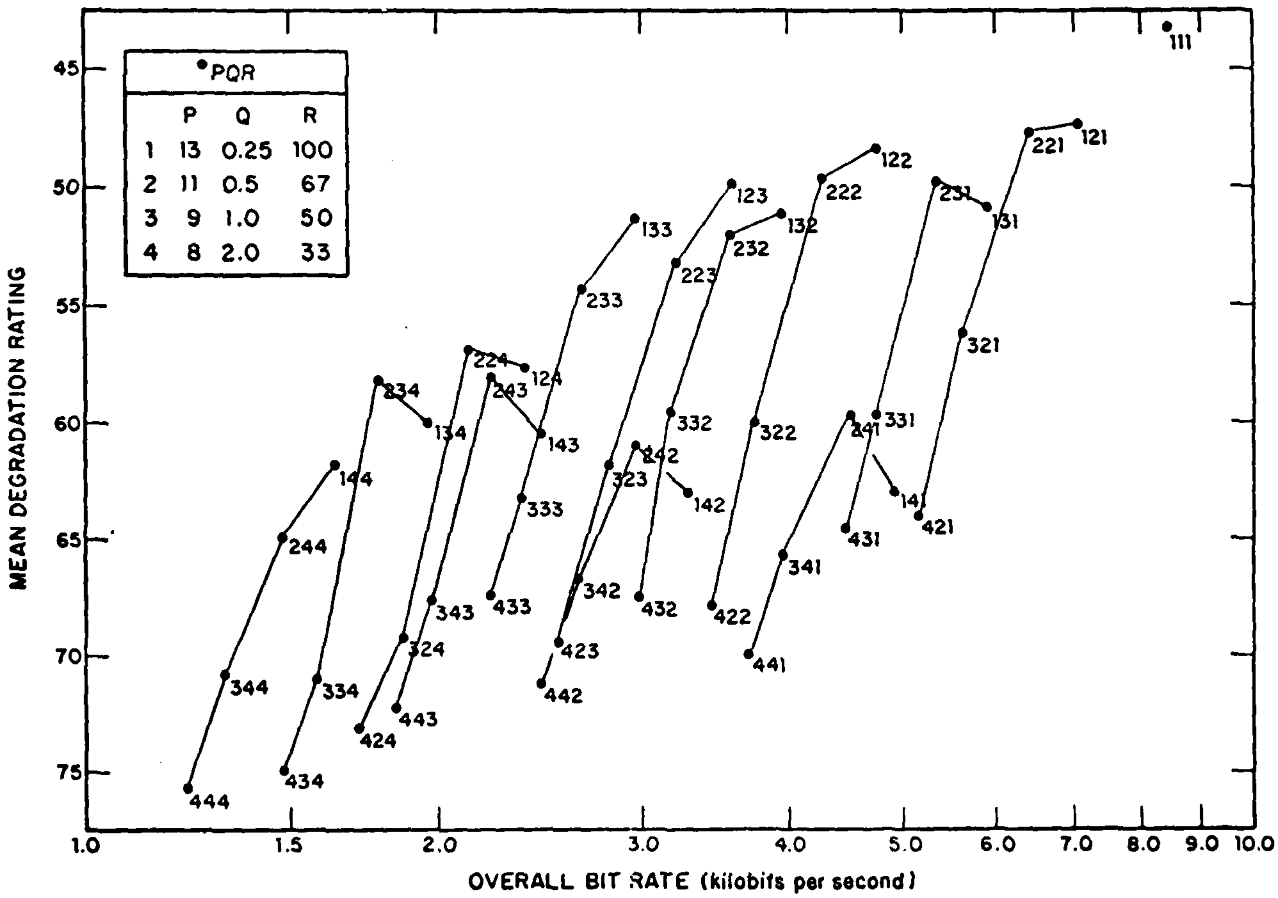

Figure 4: Degradation vs. Bit Rate. The twelve lines the rate at which degradation increases as bit rate is reduced by decreasing the number of poles from 13 to 11 to 9 to 8, with quantization step size and frame rate held constant.

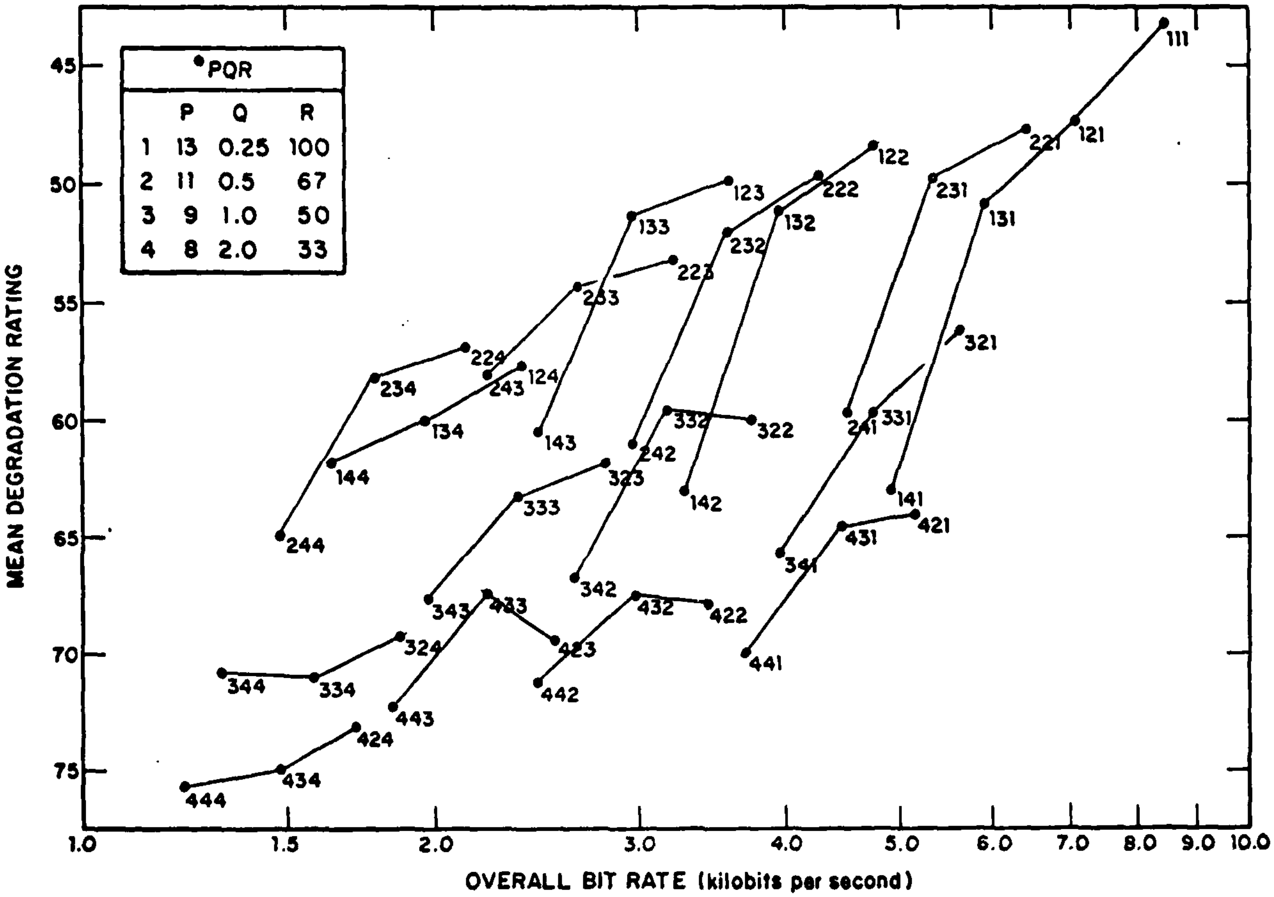

Figure 5: Degradation vs. Bit Rate. The sixteen lines show the rate at which degradation increases as bit rate is reduced by increasing the quantization step size from (0.25dB) to 0.5 dB to 1.0 dB to 2.0 dB, with number of poles and frame rate held constant.

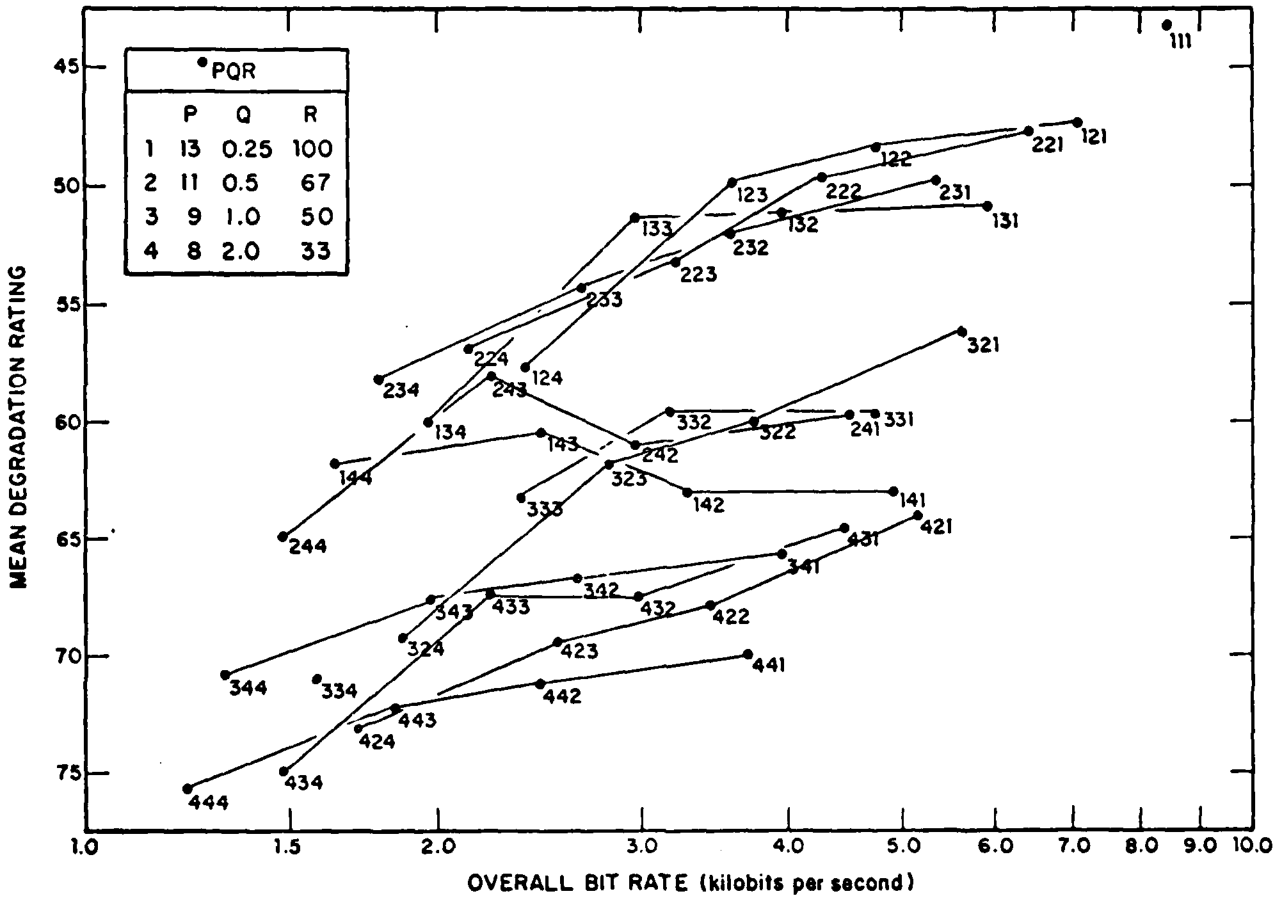

Figure 6: Degradation vs. Bit Rate. The twelve lines show the rate at which degradation increases as bit rate is reduced by decreasing the frame rate from 100/sec to 67/sec to 50/sec to 33/sec, with number of poles and quantization step size (i.e. bits per frame) held constant.

th. 11 poles is in fact the lowest number for good quality with male voices.

Further analyses of these data, including multidimensional analysis, will be reported in the next QPR.

VI. REFERENCES

1. BBN Quarterly Progress Report, Command and Control Related Computer Technology, BBN Report No. 3479, December 1976.

2. R. Viswanathan and J. Makhoul, "Specifications for ARPA-LPC System II," NSC Note 82, February 1976.

3. BBN Quarterly Progress report, Command and Control Related Computer Technology, BBN Report No. 3430, September 1976.

4. R. Viswanathan, J. Makhoul and W. Russell, "Optimal Interpolation in Linear Predictive Vocoders," BBN Report No. 3065, April 1975.

5. R. Viswanathan, "Variable Frame Rate Transmission of Pitch and Gain," NSC Note 96, September 1976.

6. L. Ehrman, "Analysis of Some Redundancy Removal Bandwidth Compression Techniques," Proc. IEEE, Vol. 55, pp. 278-287, March 1967.

7. L.D. Davisson, "Data Compression Using Straight Line Interpolation," IEEE Trans. Inf. Theory, Vol. IT-14, pp. 390-394, May 1968.

8. A.W.F. Huggins, R. Viswanathan and J. Makhoul, "Speech Quality Testing of Some Variable Frame Rate (VFR) Linear Predictive (LPC) Vocoders," ARPA NSC Note 97, February 1977.

9. R. Viswanathan, and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems." IEEE Trans. 1975,ASSP-23, p. 309-321.

10. J. Makhoul, R. Viswanathan, L. Cosell, and W. Russell, "Natural Communication with Computers." BBN Report No 2976, Vol II, 1974.

11. A.W.F. Huggins, and R.S. Nickerson, "Some effects of speech materials on vocoder quality evaluations." J. Acoust. Soc. Amer., 1975,58, S-129 (A).

12. W.P. Pachl, G.E. Urbanek, E.H. Rothauser, "Preference evaluation of a large set of vocoded speech signals." IEEE Trans. 1971,AU-19, p. 216-224.

13. A.W.F. Huggins, "The Perception of Timing in Natural Speech I: Compensation within the Syllable." Language & Speech, 1968, 11, p. 1-11.

14. E.J. Williams, "Experimental designs balanced for pairs of residual effects." Australian J. Sci. Res., 1950, A3, p. 351.

15. R.A. Crowder, and J. Morton, "Precategorical Acoustic Storage." Perception & Psychophysics, 1969, 5, p. 365-375.

16. D.N. Kalikow, K.N. Stevens, and L.L. Elliot, "Development of a Test of Speech Intelligibility in Noise Using Sentence Materials with Controlled Word Predictability." BBN Report No. 3370, 1976, to appear in J.Acoust.Soc.Amer.